



# TEXT MINING FOR SOCIETAL GOOD

**Krista Lagus**, professor, digital social science

[krista.lagus@helsinki.fi](mailto:krista.lagus@helsinki.fi)

Director & co-founder of Center for Social Data Science (CSDS)

*Visiting in course Statistical Natural Language Processing, 30.3.2021*

# CAN TEXT ANALYSIS HELP US ALLEVIATE INEQUALITY? HOW?

Social media text analytics can help us to

- increase **shared awareness**
- of various **people's experiences** of life in a society,
- as they are situated in **particular life & societal circumstances**

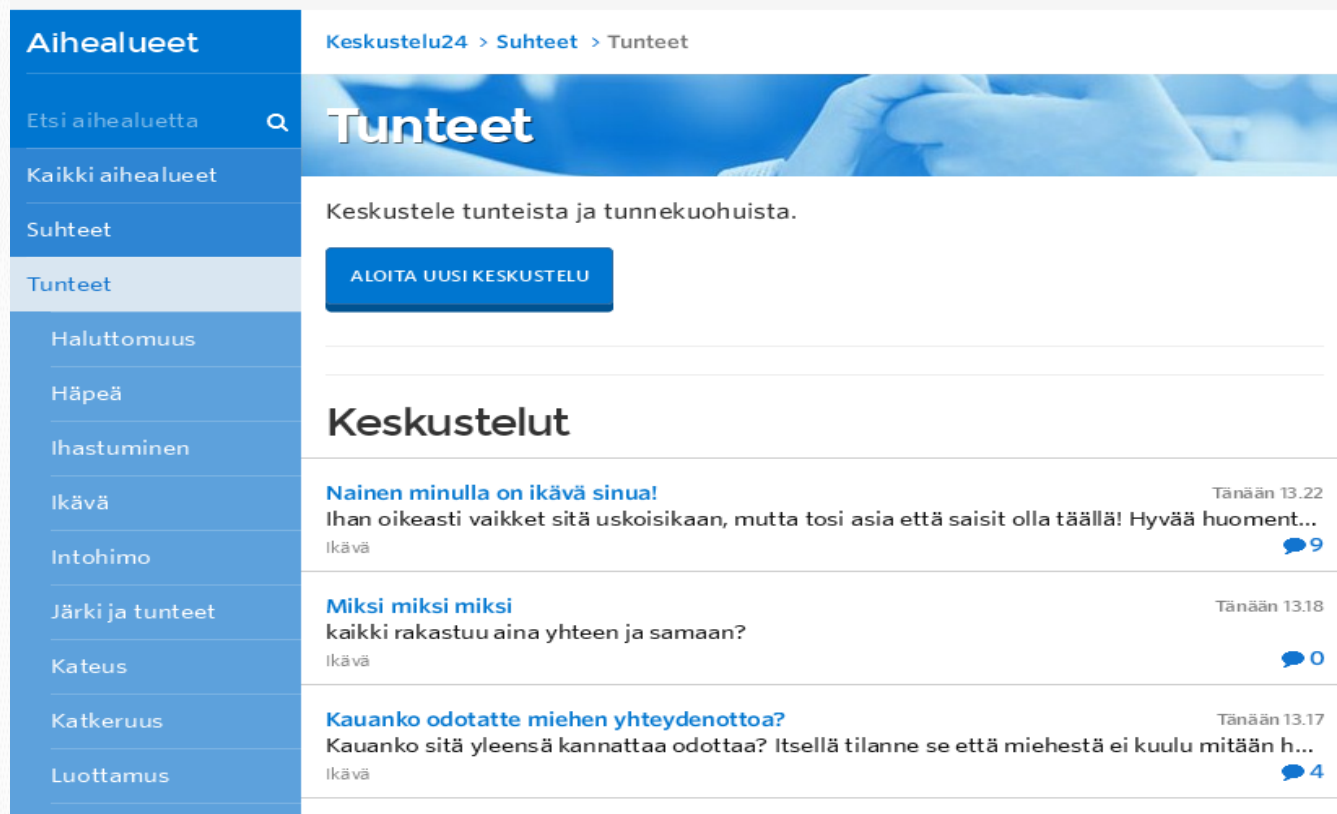
**“examination of subjective realities”** armed with compassion

Underlying assumption: What people **choose** to talk about, **matters to them** (and therefore, it is of importance to the society)

# Data: SUOMI24 forum "NATION'S BACK YARD"

80% OF FINNS USE  
MONTHLY  
OVER 80 MILLION  
COMMENTS  
FROM 17 YEARS

Owned previously  
by Aller Oy  
(Data was obtained  
from Aller to National  
Infrastructure  
Language Bank of  
Finland)



The screenshot shows the SUOMI24 forum interface. On the left is a sidebar with a search bar and a list of topic categories: Aihealueet, Etsi aihealuetta, Kaikki aihealueet, Suhteet, Tunteet, Haluttomuus, Häpeä, Ihastuminen, Ikävä, Intohimo, Järkeä ja tunteet, Kateus, Katkeruus, and Luottamus. The main content area is titled 'Tunteet' and includes a breadcrumb trail 'Keskustelu24 > Suhteet > Tunteet'. Below the title is a search bar and a button 'ALOITA UUSI KESKUSTELU'. The main content area displays a list of discussion threads under the heading 'Keskustelut'. The threads are: 'Nainen minulla on ikävä sinua!' (Tänään 13.22), 'Miksi miksi miksi' (Tänään 13.18), and 'Kauanko odotatte miehen yhteydenottoa?' (Tänään 13.17).

**Aihealueet**

Etsi aihealuetta

Kaikki aihealueet

Suhteet

**Tunteet**

Haluttomuus

Häpeä

Ihastuminen

Ikävä

Intohimo

Järkeä ja tunteet

Kateus

Katkeruus

Luottamus

Keskustelu24 > Suhteet > Tunteet

## Tunteet

Keskustele tunteista ja tunnekuohuista.

ALOITA UUSI KESKUSTELU

### Keskustelut

**Nainen minulla on ikävä sinua!** Tänään 13.22  
Ihan oikeasti vaikket sitä uskoisikaan, mutta tosi asia että saisit olla täällä! Hyvää huoment...  
Ikävä 9

**Miksi miksi miksi** Tänään 13.18  
kaikki rakastuu aina yhteen ja samaan?  
Ikävä 0

**Kauanko odotatte miehen yhteydenottoa?** Tänään 13.17  
Kauanko sitä yleensä kannattaa odottaa? Itsellä tilanne se että miehestä ei kuulu mitään h...  
Ikävä 4

A FINDING: During global financial crisis 2008-2009, the rise in social support expenditure coincides with increase in mentions of the derogatory term "Social Bum"

**Search in Korp:**

"sosiaalipummi"

or

"sossupummi"

**Result:**

over 70,000

cases

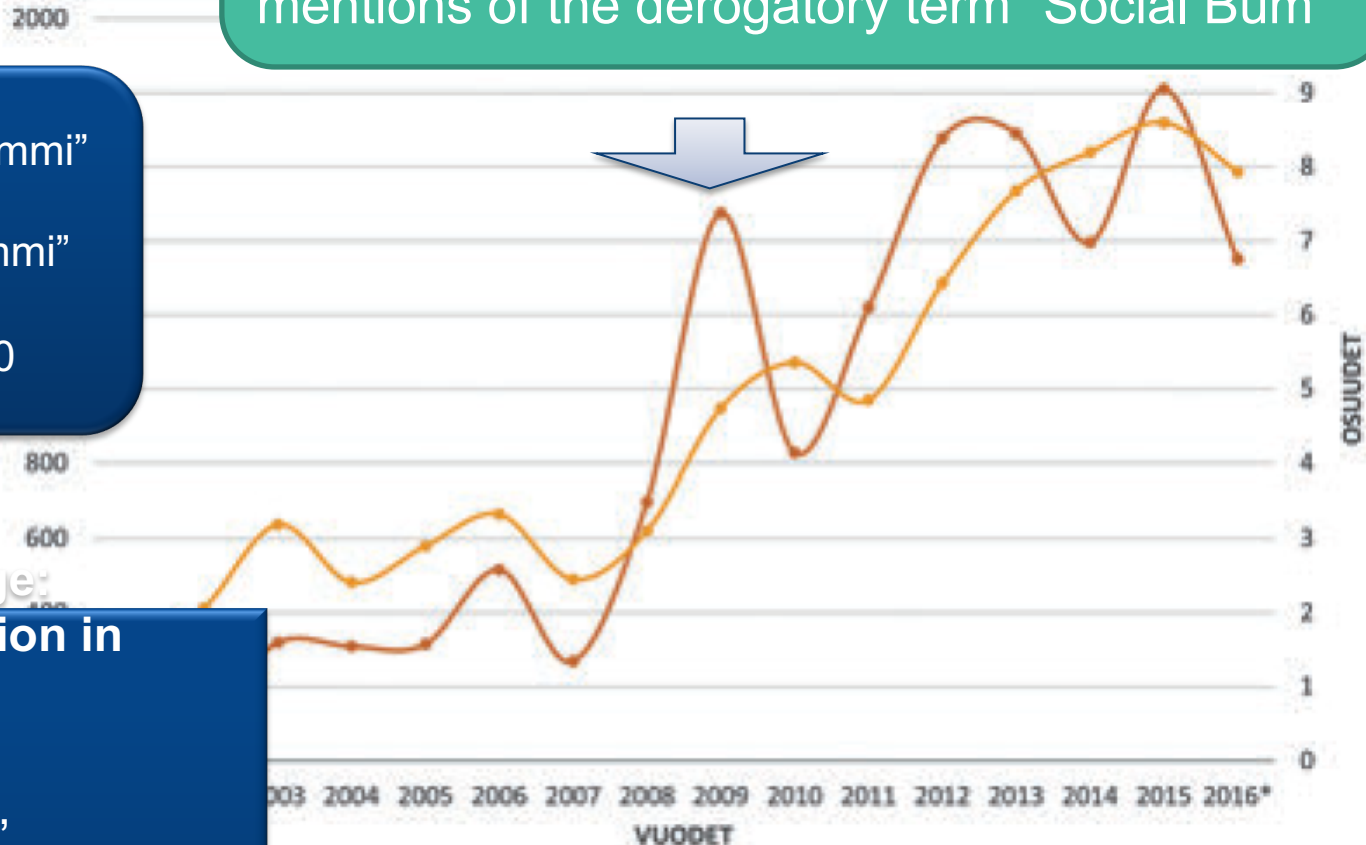
Research challenge:

**Variety of expression in Finnish**

loinenpummi,  
loinenpummiyrittäjä,  
työtönpummi, tukipummi,  
pummieliitti, ....

**Excluded:**

matkustaa pummilli,



Saari, J., Behm, M., & Lagus, K. (2017).  
Sosiaalipummi!–Moraalipaniikki 2010-luvun  
Suomessa. In Saari, J.  
(Ed.): *Sosiaaliturvariippuvuus: sosiaalipummit  
oleskeluyhteiskunnassa?*. pp. 207-232, Tampere:  
Tampere University Press.



# KEYWORD SEARCHES

... olevan yksipuolista ja se alkoi	ahdistaa	... vaan se, että epäonnistun siinä.
'kkuvippi on voinut olla joskus	Pelottaa	isokin vippi ja apu joillekin.
... asioita tietävänsä, kenellä on	hädässä	ja millainen hätä on.
... sä, kenellä on hätä ja millainen	hätä	on.
... in propagandaa ja aiheettomia	Ahdistuksen	ja vainon aika on täällä.
... attomuutta, mutta voimakasta	pelkoja	
... nä hyvin tai huonosti, ja vaikka	ahdistusta	, asenne on silti aika hälläväläi meinkinä )
... in lääkkeen käytön juuri noiden	levottomien	jalkojen takia.
... Samanlaista jalkojen	levottomuutta	ja ärsyttävää tunnetta aiheuttaa
... ta pilleriä niin ei saa nukuksi	levottomuuden	takia.
... nut hyvän arvon rasismikortin	pelossa	, tai sitten rosinalainen on vääre
... Myös vieraat koirat alkoivat	pelottaa	ja se rupesi vetämään hinnassa.
... 'kkihallissa jossa se meni aivan	paniikkiin	, tuli tulikumaksi, ei juonut vettä ja pelkäsi
... itten, ei liene tarvi pettymyksiä	tuska	niille, jotka lapsia eivät voi saada vaikka hal
... i naisilla tuntuu olevan kauhea	pelätä	
... kkea tässä maailmassa tarvitse	hätä	ja kiire suurille sanoille jne. Ja sitten niitä nä
	pelätä	!

## Approach:

1. Define concept & its textual features
2. Apply search tool: KORP
3. Observe text samples

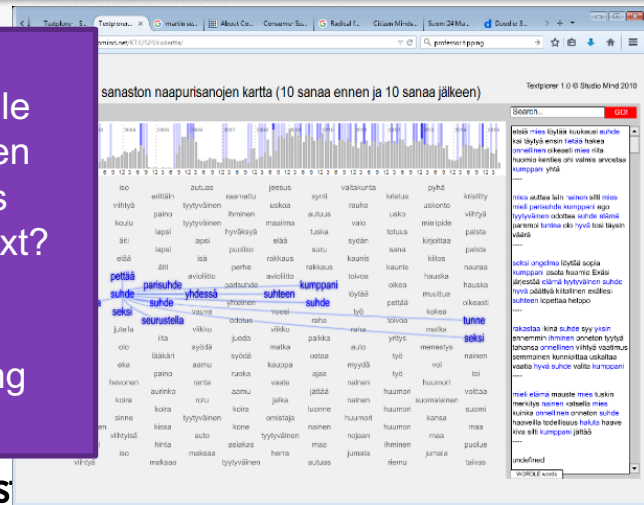


# EXPLORATORY TOPIC ANALYSIS

## Useful for:

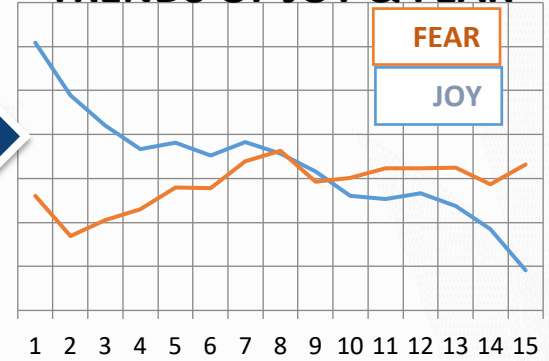
What do people talk about when fear & worry is identified in text?

## AI METHOD: Self-Organizing Map

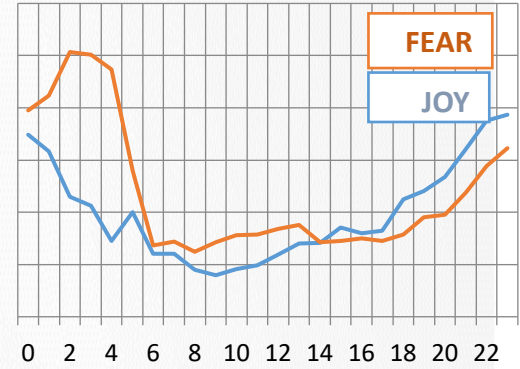


# TIME: TRENDS AND RYTHMS

## TRENDS OF JOY & FEAR



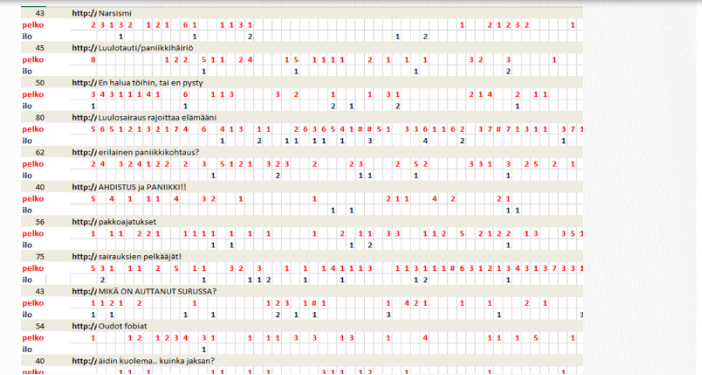
## DAILY RYTHMS



## Approach:

1. Based on existing concept definitions,
2. Calculate & visualize trends and rhythms over time

# TIME: DYNAMICS IN THREADS



## Visual observation of concept over time in individual discussion threads

# DISCUSS

## Measuring public sentiments or emotions

### Discuss with group:

1. Do you think it is possible to detect speaker's emotions from text? Explain!
2. What good might there be, if we could create a "WORRY-O-METER"?
3. What problems do you foresee?

# MEDICINE RADAR: DISCUSSIONS CONCERNING HEALTH

**Question:** How do people discuss medicines?

**Data:** Suomi24 Health discussions:

19 million comments, 200 million words, 16 years

**Challenge:** *Can concepts of interest be found using AI, augmented with human input?*

## Hybrid method

- Word embeddings
- Linguistic analyses
- Human feedback
- Statistical relevance measure

**Collaboration:** Citizen Mindscapes research project & Futurice company

### Research article:

K. Lagus, M. Ruckenstein, A. Juvonen, C. Rajani, Digital Humanities Norden, 2018

**Software:** published open-source

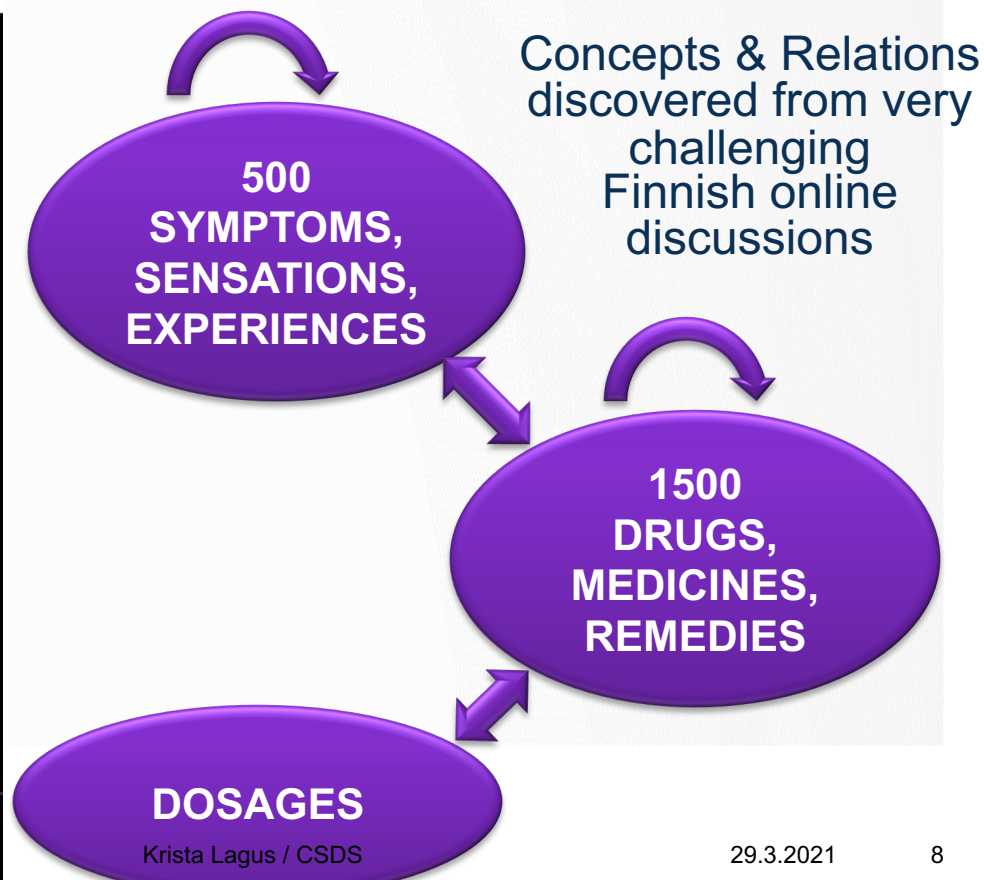
# MEDICINE RADAR: WHAT WE ACHIEVED

## Handling linguistic variation

### Words interpreted as unettomuus ×

unettomuudestakin, unettomuuden, unettomuutena, unettomuuteesi, unettomuutta, unettomuudelle, unettomuuteeni, unettomuudelta, unettomuuksista, unettomuuteenhan, unettomuuteenkin, unettomuudet, unettomuudeksi, unettomuutenne, unettomuutensa, unettomuuteenko, unettomuus-, unettomuuttakin, unettomuutesi, unettomuuskin, unettomuudestani, unettomuudesta, unettomuudetkin, unettomuuksissa, unettomuudessa, unettomuuttani, unettomuushan, unettomuudenkin, unettomuuttanne, unettom, unettomuuttasi, unettomuudestakaan, unettomuudestanne, unettomuuksiin, unettomuudeks, unettomuuteni, unettomuuttanikin, unettomuudeltani, unettomuudestaan, unettomuus, unettomuudessani, unettomuuskaan, unettomuuttaan, unettomuudestanikaan, unettomuuttakaan, unettomuuksia, unettomuuteenkaan, unettomuuteen, unettomuuksineen, unettomuuteenne, unettomuudella, unettomuuden-, unettomuudessasi, unettomuudestasi, unettomuuteensa

## Discovering proto-ontology for medicines and symptoms





# MEDICINE RADAR search & exploration tool: Find SYMPTOMS related to "BURANA" (a painkiller)

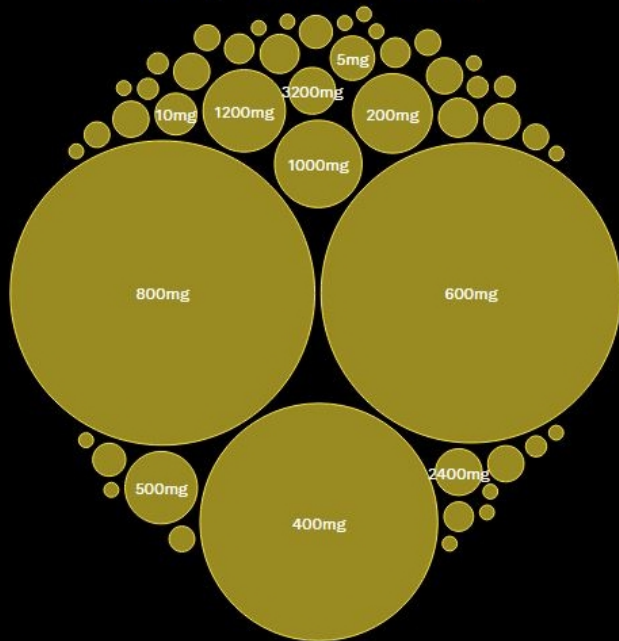
Lääketutka

burana

SEARCH

## Related Dosages

Dosages mentioned with burana



! This is not medical advice, the data only reflects what people talk about.

## Related Symptoms

Relevant terms mentioned with burana

hammassärky  
kuukautiskipu  
selkäsärky  
jomottaa  
hermokipu  
hermosärky  
särkeä  
pääkipu  
iskias  
angiina  
kipu  
kurkkukipu

COULD SIMILAR BE  
DONE  
FOR SOCIAL & HEALTH  
CUSTOMERS &  
PROFESSIONALS  
TALK?

Project: Citizen Mindscapes

Article:

K. Lagus, M. Ruckenstein, A. Juvonen, C. Rajani, "Medicine Radar – A tool for exploring online health discussions." Digital Humanities Norden, 2018

<http://www.laaketutka.fi>

! This is

# DISCUSS

Try out Medicine Radar at: [Laaketutka.fi](https://laaketutka.fi)

E.g.

- search for a medicine you know
- Look at the list of other medicines, other symptoms, typical dosages. Is this what you would expect?

## Discuss:

1. What is interesting about this prototype?
2. What could it be useful for?
3. Come up with questions & share them with your group!

AI, ROBOTICS & PLATFORMS

REDISTRIBUTION OF WORK

REDISTRIBUTION OF DATA:  
MYDATA, GDPR

REDISTRIBUTION OF POWER

BLOCK CHAIN

CRYPTO CURRENCIES

DIGITAL CONSUMERISM  
IN GLOBAL MARKETS  
& LOCAL WORK CONDITIONS

# SOCIETIES ARE UNDER RAPID SYSTEMIC CHANGE

PANDEMIA

DIGITALIZATION OF  
COMMUNICATION

CLIMATE & NATURAL  
ENVIRONMENT

ALTERNATIVE FACTS, BUBBLES,  
CONFLICTING WORLD VIEWS

ELDERLY SOCIETIES

NATIONS ON THE MOVE

**ARE SOCIAL SCIENCES  
KEEPING UP WITH THE PACE?**



# WORK IN PROGRESS CHANGING PERSPECTIVE

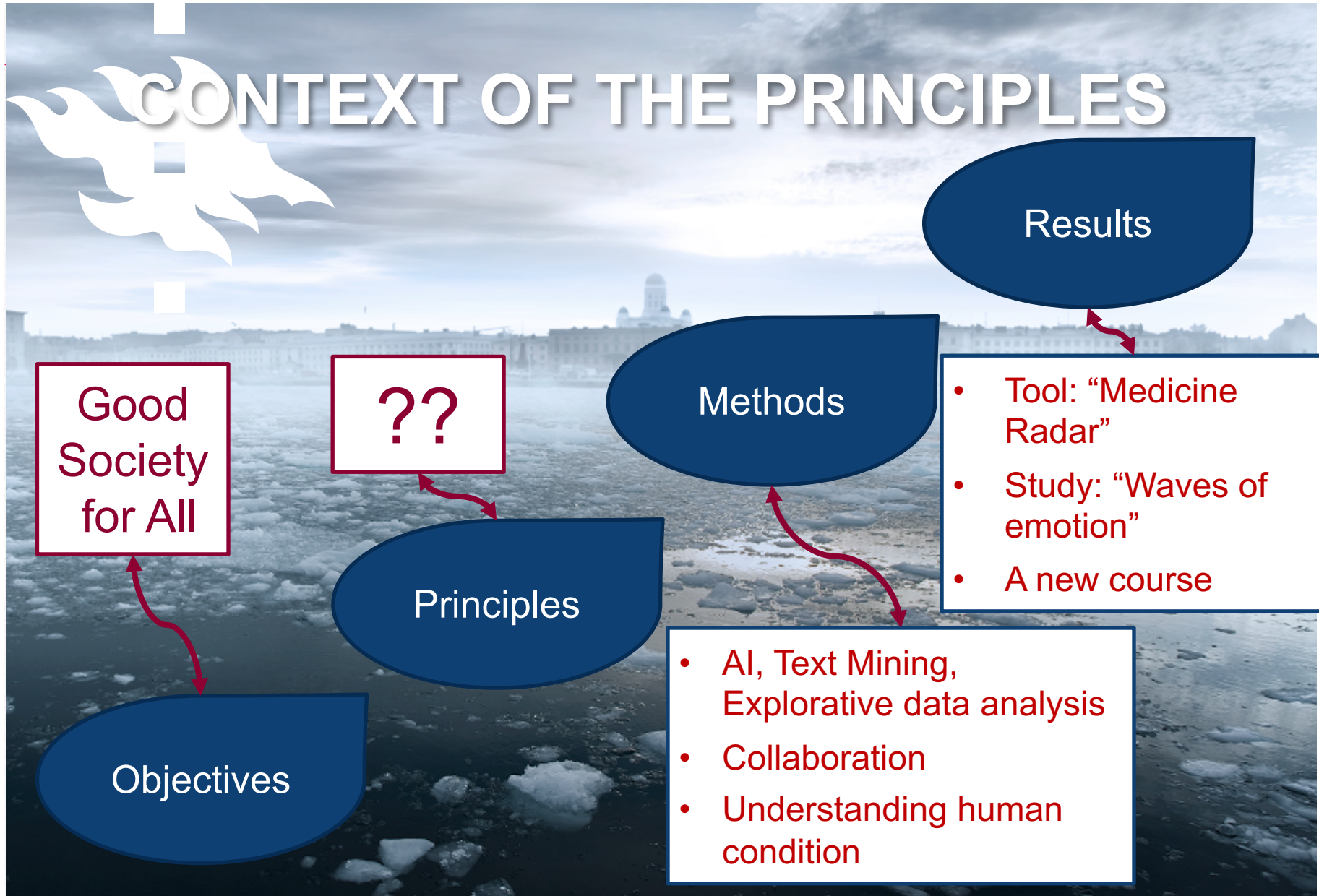
FROM (RESEARCH)  
OUTCOMES

TO UNDERLYING  
PRINCIPLES





# CONTEXT OF THE PRINCIPLES





# MEDICINE RADAR & ITS PRINCIPLES

## What we refrained from:

- **No Fixed truth perspective:**  
instead of starting with an agreed list of medicines, a “medicine” was considered **anything that people talked about as if a medicine or remedy**

E.g. **kannabis** was allowed as a “medicinal” concept since it was apparently used in that way in the discussions

## Principles in use:

- **Compassionate awareness**  
Just observing the conversations without judgement
- **Empowering & facilitating** the user by concept-based exploration approach
- **Valuing variety of perspectives**  
everyone’s story was considered *equally valuable*
- **Collaboration & cocreation**
  - btw data science & social sciences
  - btw university & private sector
  - btw researchers & data analyst & programmer & web designer

# DIRECTIONS: WHAT *PRINCIPLES* TO CONSIDER IMPLEMENTING IN THE DIGITAL

## FROM OVER-USE OF POWER TO EMPOWERING EVERYONE

- From *power over* to *power with others* (adopted from Terhi Takanen, 2012)
- From opposing action to joining with it (*a budo principle*)
- Collaboratively forming the final direction
  - *Participatory processes*
  - *Collaborative work methods*
  - *Co-creating with others*
  - *Appreciative dialogue*

New course:  
CoCreative methods of  
collaboration in Societal  
challenges



# A CHALLENGE: PANDEMIA & BUBBLES

## How to recognize and agree on what is a fact?

- E.g. does COVID19 exist?
- What number of cases are there right now?
- Which prevention measures are effective?
- What of the adverse effects of the measures?

## How to *clarify understanding* of societal processes to everyone?

- E.g. **How is registry data collected about covid cases?**
- Context: THL is facing critique & questions & is probably doing their best to answer
- **Our own early progress: A new Registry data course by CSDS ongoing (by Maria Valaste)**

**AS A SOCIETY, WE ARE NOW RE-NEGOTIATING EVERYTHING AT ONCE**



# PRINCIPLES.... VALUING EVERYONE

## VALUING OTHERS & OURSELVES:

Do we avoid or embrace variety?

- Do we apply **Respectful dialogue** (E.g. **Erätauko** method employed by Sitra)
- Do we empirically study & bring to view the **words, concepts & expressions that different people use** (without judgement)?
- Do we take care that **our data is representative of ALL**, or **of OURSELVES & OUR KIN**?
  - Gender: Male/Female/other
  - Age: Kids, Youth, Young adult, Mature, Old
  - Socio-economic: The “hard-to-find” groups: those without a phone, those without a home, the addicted etc (not just the capable, aware, most intelligent, well-to do people)

# PRINCIPLES... AWARENESS, CONNECTEDNESS

## FOSTERING AWARENESS

- *Reflexive practices* applied for understanding “self” and “us”
  - Avoidance on focusing on “them”, the “others”, and “what is wrong with them”
  - *How can we facilitate self-reflection in digital platforms?*

## FOSTERING CONNECTEDNESS

- Are we willing to move from asking what separates us *to asking what connects us?*
  - *What common grounds can we (help everyone to) identify?*
  - *How can we (provide everyone the) experience of connection to others & the nature?*

# PRINCIPLES... ON ADDRESSING IGNORANCE

## **DIMINISHING IGNORANCE:**

WHICH STRATEGY DO WE EMPLOY?

- **Do we tell others the “correct facts” & employ everyone to watch their neighbour and report “to the system” everyone who tells “incorrect facts”**  
“STASI APPROACH”: fixed mindset + stationary world view + negative peer culture
- **Do we build processes that ask questions and help everyone in asking more questions, in order to develop everyone’s thinking**  
“SOCRATIC APPROACH”: growth mindset + open-ended/dynamic world view + dialogue/collaborative culture

## **FROM FEAR CULTURE TO TRUST CULTURE**



# PRINCIPLES ... POSITIVE & BALANCED RATHER THAN HYPER CRITICAL

- **Recognize tendency** to focus on bad, unwell or wrong.
- **Resist the lure of joining the fight against the “bad”** in self, others and group & social level
- **Practice joining with whatever is good**
- **How: By identifying & strengthening good:** recognizing, seeing, mentioning, rewarding by attention, resourcing, helping others recognize in themselves, in others, and in societal practices

## STRATEGY:

Melting bubbles with love, compassion, awareness and  
by building safety and trust



# DISCUSS

1. What are important principles for you, that you would like to see more of, in the world / in the discussions / in social media?
  - Write down at least one, and describe it to your group
2. What would the world be like if your chosen principle was adopted or became stronger in the world, or in some particular context or forum? Describe concretely, if possible



# RELATED PUBLICATIONS

- Mäkelä, E., **Lagus, K.**, Lahti, L., Säily, T., Tolonen, M., Hämäläinen, M., ... & Nevalainen, T. (2020). Wrangling with non-standard data. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference Riga, Latvia, October 21-23, 2020*. CEUR-WS.org.
- **Lagus, K.**, Pantzar, M. & Ruckenstein, M. (2018). Kansallisen tunnemaismen rakentuminen: Pelon ja ilon rytmit verkkokeskusteluissa. *Kulutustutkimus.Nyt*. 12:1-2, pp. 62-83
- **Lagus, K. H.**, Ruckenstein, M. S., Juvonen, A. & Rajani, C., 2018, Medicine Radar – a Tool for Exploring Online Health Discussions. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference*. Mäkelä, E., Tolonen, M. & Tuominen, J. (eds.). Aachen: CEUR-WS.org, p. 460-468 8 p.
- Saari, J., Behm, M., & **Lagus, K.** (2017). Sosiaalipummi!–Moraalipaniikki 2010-luvun Suomessa. In Saari, J. (Ed.): *Sosiaaliturvariippuvuus: sosiaalipummit oleskeluyhteiskunnassa?*. pp. 207-232, Tampere: Tampere University Press.
- **Lagus, Krista**, Pantzar, Mika, Ruckenstein, Minna ja Ylisiurua, Marjoriikka (2016). [Suomi 24 – Muodonantoa aineistolle](#). Kuluttajatutkimuskeskus, Valtiotieteellisen
- Ruckenstein, M. & Turunen, L., (2019). Re-humanizing the platform: Content moderators and the logic of care. (Accepted/In press) In : *New Media & Society*.
- Ruckenstein, M. (2019). Tracing medicinal agencies: Antidepressants and life-effects. In: *Social Science & Medicine*.
- Suominen, Jaakko; Saarikoski, Petri; Vaahensalo, Elina (2019): Digitaalisia kohtaamisia. Verkkokeskustelut BBS-purkeista sosiaaliseen mediaan [Digital encounterings. Online discussions from bulletin board systems to social media]. Helsinki: Gaudeamus.
- Pink, S., Ruckenstein, M. S., Willim, R. & Duque, M., (2018). Broken data: Conceptualising data in an emerging world; *Big Data & Society* 5:1
- Harju, Auli. (2018). Suomi24-keskustelut kohtaamisten ja törmäysten tilana. *Media & viestintä* 41(1), 51-74.
- Vaahensalo, Elina (2018). ”Keskustelufoorumit mediainhokkeina – suositut suomenkieliset keskustelufoorumit mediassa.” *Wider screen*, 3, 2018
- Vaahensalo, Elina (2017). ”Kaikenkattavista portaaleista anarkistiseen sananvapauteen. Suomalaisten verkkokeskustelufoorumien vuosikymmenet.” *Tekniikan Waiheita* 2/2017.
- Östman, S., & Turtiainen, R. (2016). From research ethics to researching ethics in an online specific context. *Media and Communication*, 4(4), 66-74. ([Available as open access](#))
- Suominen, Jaakko & Sivula, Anna (2016). Digisyntyisten ilmiöiden historiantutkimus. Teoksessa: Elo, Kimmo (toim.). *Digitaalinen humanismi ja historiatieteet*. Turku: Turun historiallinen yhdistys.



**THANK YOU!**

[Krista.lagus@helsinki.fi](mailto:Krista.lagus@helsinki.fi)

